
intake_mongo Documentation

Release 0.1.0

Joseph Crail

Feb 18, 2020

Contents:

1 Quickstart	3
1.1 Installation	3
1.2 Usage	3
2 API Reference	5
3 Indices and tables	7

This package allows loading MongoDB data into python using Intake.

`intake-mongo` provides quick and easy access to tabular data stored in [MongoDB](#)

This plugin reads MongoDB collections without random access: there is only ever a single partition.

1.1 Installation

To use this plugin for `intake`, install with the following command:

```
conda install -c intake intake-mongo
```

1.2 Usage

1.2.1 Ad-hoc

After installation, the function `intake.open_mongo` will become available. It can be used to fetch a collection from the MongoDB server, and download the results as a list of dictionaries.

The parameters are of interest when defining a data source:

- `uri`: a string like `'mongodb://localhost:27017'` to reach the server on. Additional connection information can be supplied as part of the URI or with the separate `connection_kwargs` parameter (a key-value dictionary).
- `db`: the name of the mongo database to access
- `collection`: a string like `'test_collection'` identifying a dataset on the server, within the given database
- `find_kwargs`: a broad range of possible parameters to pass to the `find` method, including filtering, sorting, choosing of fields

1.2.2 Creating Catalog Entries

To include in a catalog, the plugin must be listed in the plugins of the catalog:

```
plugins:
  source:
    - module: intake_mongo
```

and entries must specify `driver: mongo`.

1.2.3 Using a Catalog

A full entry might look like:

```
sources:
  sample1:
    driver: mongo
    args:
      uri: "mongodb://localhost:27017"
      db: "test-database"
      collection: mycollection
      connect_kwargs: {"ssl": true}
      find_kwargs: {'projection': ['field1', 'field2']}
      _id: false
```

In this case, we specify a connection to the local machine, connect with SSL activated, and select only “field1” and “field2” of `mycollection` to retrieve, with the default ID column not included in the output.

Note that the set of options available will depend on the version of `pymongo` installed.

CHAPTER 2

API Reference

```
intake_mongo.intake_mongo.  
MongoDBSource
```

CHAPTER 3

Indices and tables

- `genindex`
- `modindex`
- `search`